

On Stochastic Estimation of Partition Function

Ali Al-Bashabsheh Yongyi Mao
 School of Electrical Engineering and Computer Science
 University of Ottawa, Canada
 {aalba059, yymao}@eecs.uottawa.ca

Abstract—In this paper, we show analytically that the duality of normal factor graphs (NFG) can facilitate stochastic estimation of partition functions. In particular, our analysis suggests that for the q -ary two-dimensional nearest-neighbor Potts model, sampling from the primal NFG of the model and sampling from its dual exhibit opposite behaviours with respect to the temperature of the model. For high-temperature models, sampling from the primal NFG gives rise to better estimators whereas for low-temperature models, sampling from the dual gives rise to better estimators. This analysis is validated by experiments.

I. INTRODUCTION

The estimation of partition function for statistical models is of fundamental importance in statistical physics, machine learning and information theory [1], [2]. The models we consider in this paper are specified by a collection of random variables $\{X_i : i = 1, 2, \dots, N\}$, for some positive integer N ; each random variable X_i is assumed to take values (often called *spins*) from some finite set \mathcal{X} ; every configuration $x \in \mathcal{X}^N$ is associated with an energy level $E(x)$, and the joint distribution of random variables $\{X_i : i = 1, 2, \dots, N\}$ is modelled as the Boltzmann distribution

$$p_B(x) := \frac{e^{-\beta E(x)}}{Z}, \quad (1)$$

for all $x \in \mathcal{X}^N$. In (1), $\beta := \frac{1}{kT}$ is often referred to as the “inverse temperature”, where T is the temperature and k is the Boltzmann constant, and the normalizing constant $Z := \sum_{x \in \mathcal{X}^N} e^{-\beta E(x)}$ is known as the *partition function*.

Given β and the energy function $E(\cdot)$, exact computation of the partition function Z for systems involving a large number of random variables is known to be intractable, and it is precisely the intractability of this problem that roots the hardness of various problems in coding and information theory (e.g., determining the capacity of constrained codes). Developing bounding techniques (e.g. [1]) and approximation methods [3] for estimating the partition functions is thus an active area of research.

This work is motivated by the recent empirical observation of [4] that for the two-dimensional nearest-neighbor Ising model (binary spins), the duality of normal factor graphs (NFG) [5] appears to facilitate the estimation of the partition function. In particular, they experimentally show that for large β , two stochastic estimation methods (the Ogata-Tanemura method [6] based on Gibbs sampling and a method based on uniform sampling) provide better estimation of the partition function when sampling from the dual NFG compared to sampling from the primal NFG.

In this paper, we explain the behaviour observed in [4] and show both analytically and experimentally that such a trend extends beyond the Ising model to q -ary spins, i.e., the standard Potts model [7]. Along our development, we also provide insights on the question for what other two-dimensional nearest-neighbor models such behaviour may hold.

The remainder of this paper is organized as follows. In Section II, we precisely state the model considered in this paper and present the NFG representation of the model and the duality result therein [5]. A concise review of two stochastic estimation methods and the approach of [4] are also given in Section II. Section III presents an analysis of the convergence behaviour of the two methods, suggesting that in high β regime, sampling from the dual NFG model performs better whereas in low β regime, sampling primal NFG performs better. The analysis is supported by the experimental results presented in Section IV. The paper is concluded in Section V, where we extend the results beyond the Potts model.

II. PRELIMINARIES

A. Model

In Equation (1), we consider that each index in $\{1, 2, \dots, N\}$ corresponds to a grid point in an $L \times L$ square lattice. We assume that the lattice is “wrapped around” in the sense that the left-most point of each row is connected to the right-most point of the same row and the top-most point of each column is connected to the bottom-most point of the same column. Let \mathcal{A} denote the set of all pairs of adjacent lattice points. The energy function is assumed to take the form

$$E(x) := - \sum_{\{i,j\} \in \mathcal{A}} g_{ij}(x_i, x_j), \quad (2)$$

for a collection of functions $\{g_{ij} : (i, j) \in \mathcal{A}\}$. Such a model is referred to as a two-dimensional nearest-neighbor model.

We will further assume that the alphabet \mathcal{X} is the abelian group $\mathbb{Z}_q := \{0, \dots, q-1\}$ and that

$$g_{ij}(x, x') = g(x, x') := \begin{cases} 1, & x = x' \\ -1, & x \neq x'. \end{cases} \quad (3)$$

Equations (1) to (3) define a (two-dimensional nearest-neighbor) Potts model.¹ (Some authors use the term *standard* Potts model to make explicit the distinction from the “clock”

¹We slightly deviate from the traditional definition of the Potts model where the function g is usually assumed to take the value 0 instead of -1 . Without altering the nature of the problem, this choice of function g includes the Ising model as the special case of $q = 2$.

model.) To facilitate later discussions, we use f_B to denote $e^{-\beta E(x)}$ in (1) and refer to it as the “unnormalized Boltzmann distribution”.

B. NFG Representation and Duality

A normal factor graph (NFG) \mathcal{G} is a graph $(\mathcal{V}, \mathcal{E})$ where each edge $e \in \mathcal{E}$ is associated a variable x_e , and each vertex $v \in \mathcal{V}$ is associated a local function $f_v(x_{E(v)})$, where $E(v)$ is the set of edges incident with v , and for any set \mathcal{A} , $x_{\mathcal{A}} := \{x_a : a \in \mathcal{A}\}$. Let $\mathcal{X}_{\mathcal{G}}$ be the support of the function defined as the multiplication of all local functions, and let $f_{\mathcal{G}}$ be the restriction of such function to $\mathcal{X}_{\mathcal{G}}$. Further, we define $Z_{\mathcal{G}}$ as the sum of $f_{\mathcal{G}}$ over $\mathcal{X}_{\mathcal{G}}$, and write $p_{\mathcal{G}} := f_{\mathcal{G}}/Z_{\mathcal{G}}$. Note that if all the local functions are nonnegative, then $p_{\mathcal{G}}$ is a probability distribution over $\mathcal{X}_{\mathcal{G}}$. In this case, in alignment with the previous discussions, we refer to $p_{\mathcal{G}}$, $f_{\mathcal{G}}$, and $Z_{\mathcal{G}}$ as the distribution, unnormalized distribution, and partition function of the NFG, respectively. We note that the above definitions of NFG and related terms deviate slightly from those in [5]. This is to simplify our presentation and exclude the concepts irrelevant to this paper.

It is natural to associate with the model defined in Section II-A an NFG as in Fig. 1 (wrapping around is not shown). In the figure, each function marked by “=” is an “equality indicator function”, namely, a function that evaluates to 1 if all its arguments are equal and evaluates to 0 otherwise; each equality indicator function corresponds to a random variable in the model. The function h in the figure is defined by $h(x, x') := e^{\beta g(x, x')}$. It is not hard to see that the unnormalized distribution, distribution and partition functions associated with this NFG are respectively f_B , p_B and Z of the model defined by equations (1), (2) and (3).

Noting that function g only depends on the difference between its arguments, we may express h by $h(x, x') := \kappa(x - x')$, where

$$\kappa(x) = \begin{cases} e^{\beta}, & x = 0 \\ e^{-\beta}, & x \neq 0. \end{cases} \quad (4)$$

Using function κ , the NFG in Fig. 1 may be converted to the NFG in Fig. 2 without changing its unnormalized distribution, distribution and partition function. This latter NFG, which we denote by \mathcal{G} is in fact preferred in the context of this paper, since the results of this paper depend crucially on a property of κ , which will become clear momentarily.

It is possible to introduce duality to NFG via the Fourier transform. Briefly, the Fourier transform of any function f on \mathbb{Z}_q^m is another function \hat{f} defined on \mathbb{Z}_q^m . In particular, the Fourier transform of an equality indicator function is, up to scale, a “parity-check” indicator function, namely a function that evaluates to 1 if its argument sums to 0 and evaluates to 0 otherwise. A parity-check indicator function is marked by “+” in an NFG. Further, the Fourier transform of the function κ is

$$\hat{\kappa}(x) = \begin{cases} e^{\beta} + (q-1)e^{-\beta}, & x = 0 \\ e^{\beta} - e^{-\beta}, & x \neq 0. \end{cases} \quad (5)$$

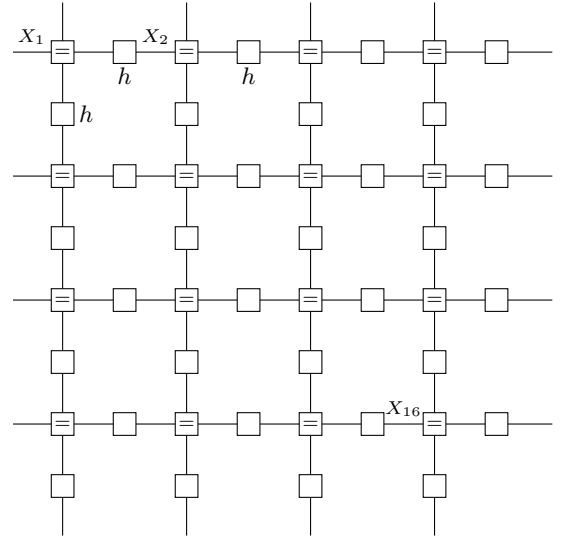


Fig. 1. An NFG representing the model specified by (1), (2) and (3).

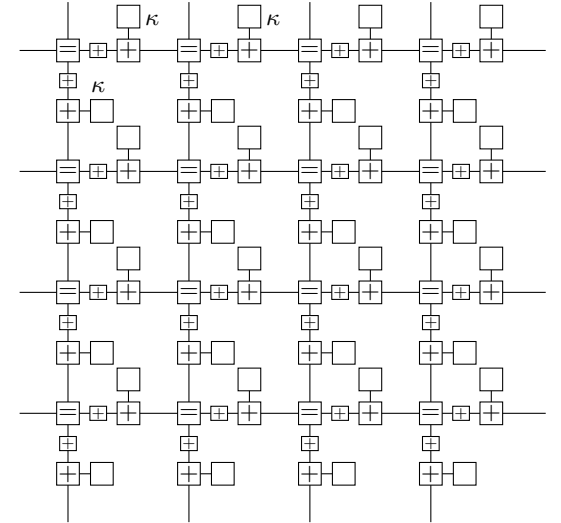


Fig. 2. The NFG \mathcal{G} .

Given an NFG, the dual NFG may be obtained by converting each local function to its Fourier transform and then inserting a parity-check indicator function to each edge. It can then be verified that the dual NFG of \mathcal{G} is the NFG \mathcal{G}' shown in Fig. 3. A duality theorem (generalized Holant theorem) of NFG [5] states, in the context of our model, that $Z_{\mathcal{G}'} = Z_{\mathcal{G}}/q^N$.

C. Estimating Partition Function by Sampling NFG

Given an NFG \mathcal{G} representing a statistical model, its partition function $Z_{\mathcal{G}}$ may be estimated via evaluating its unnormalized distribution $f_{\mathcal{G}}$ at a set of configurations Y_1, Y_2, \dots, Y_M randomly drawn from $\mathcal{X}_{\mathcal{G}}$. If these configurations are obtained by sampling the distribution $p_{\mathcal{G}}$ (which in practice can be done by Gibbs sampling), then the Ogata-Tanemura (OT) [6]

estimator can be defined as

$$Z_G^{\text{OT}}(M) := \frac{|\mathcal{X}_G|}{\frac{1}{M} \sum_{i=1}^M \frac{1}{f_G(Y_i)}}, \quad (6)$$

If these samples are drawn uniformly from \mathcal{X}_G , an estimator, which we call the “uniform estimator”, can be defined as

$$Z_G^{\text{U}}(M) := \frac{|\mathcal{X}_G|}{M} \sum_{i=1}^M f_G(Y_i), \quad (7)$$

It can be shown that as M increases, both $Z_G^{\text{OT}}(M)$ and $Z_G^{\text{U}}(M)$ converges to Z_G .

Given the NFG \mathcal{G} in Fig. 2 that represents the Potts model, it is easy to see that every local function in the dual NFG \mathcal{G}' in Fig. 3 is non-negative. The dual NFG \mathcal{G}' may then be regarded also as a statistical model and the above two estimators may be used to estimate the partition function $Z_{\mathcal{G}'}$, a scaled version of Z_G . This technique was first used in [4] for Ising model (Potts model with $q = 2$), where the authors show empirically that at high temperature, both OT estimator and uniform estimator give more accurate estimates on the dual NFG.

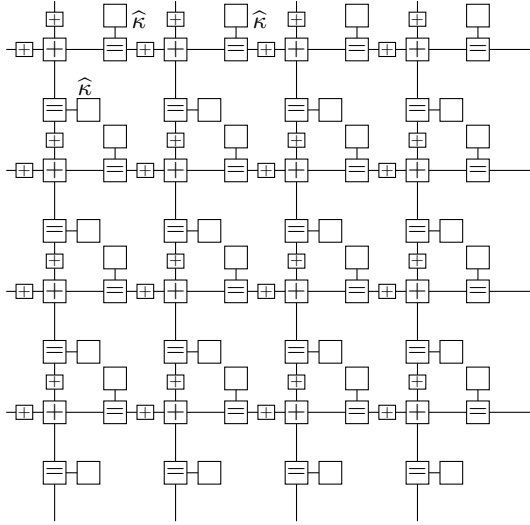


Fig. 3. The dual NFG \mathcal{G}' .

III. CONVERGENCE BEHAVIOUR OF THE ESTIMATORS

Our analysis is primarily based on bounding the respective variances of the logarithm of the estimators for large M , as for any given number M of samples, such variance is an indicator of the estimation accuracy. Our development is largely in line with that of [3].

Given a statistical model NFG \mathcal{G} , it is possible to show

$$\begin{aligned} \lim_{M \rightarrow \infty} M \text{Var}[\log(Z_G^{\text{OT}}(M))] &= \frac{Z_G^2}{|\mathcal{X}_G|^2} \text{Var}\left[\frac{1}{f_G(Y_1)}\right] \\ &= \frac{Z_G}{|\mathcal{X}_G|^2} \sum_{x \in \mathcal{X}_G} \frac{1}{f_G(x)} - 1. \end{aligned} \quad (8)$$

Proof: Let

$$X_M := \frac{\frac{1}{M} \sum_{i=1}^M \frac{1}{f_G(Y_i)}}{|\mathcal{X}_G|},$$

then

$$E[X_M] = \frac{1}{|\mathcal{X}_G|} E\left[\frac{1}{f_G(Y_1)}\right] = \frac{1}{|\mathcal{X}_G|} \sum_x \frac{p_G(x)}{f_G(x)} = \frac{1}{Z_G},$$

and

$$\text{Var}[X_M] = \frac{1}{M|\mathcal{X}_G|^2} \text{Var}\left[\frac{1}{f_G(Y_1)}\right].$$

From (6) we can rewrite $\log(Z_G^{\text{OT}})$ as

$$\log(Z_G^{\text{OT}}) = g(X_M),$$

where $g(x) := \log(\frac{1}{x}) = -\log(x)$, and so $g'(x) = \frac{-1}{x}$. Using Taylor expansion of g at $E[X_M]$,

$$\log(Z_G^{\text{OT}}) \simeq g(E(X_M)) + g'(E(X_M))(X_M - E(X_M)),$$

and so

$$\begin{aligned} \text{Var}[\log(Z_G^{\text{OT}})] &\simeq (g'(E(X_M)))^2 \text{Var}[X_M] \\ &= \frac{1}{(E[X_M])^2} \text{Var}[X_M] \\ &= \frac{Z_G^2}{M|\mathcal{X}_G|^2} \text{Var}\left[\frac{1}{f_G(Y_1)}\right] \end{aligned}$$

The approximation is only valid in the limit, as g may be approximated as a linear function only when the variance of X_M is small. This method of first order approximation is often referred to as the *delta method*. For a more rigorous discussion on the delta method, see e.g. [8, Theorem 5.5.24]. ■

Similarly, it can be shown that

$$\begin{aligned} \lim_{M \rightarrow \infty} M \text{Var}[\log(Z_G^{\text{U}}(M))] &= \frac{|\mathcal{X}_G|^2}{Z_G^2} \text{Var}[f_G(Y_1)] \\ &= \frac{|\mathcal{X}_G|}{Z_G^2} \sum_{x \in \mathcal{X}_G} f_G^2(x) - 1. \end{aligned} \quad (9)$$

From this, the following proposition can be proved.

Proposition 1. When sampling the NFG \mathcal{G} of the Potts model,

$$L_{\text{OT}}(\beta) \leq \lim_{M \rightarrow \infty} M \text{Var}[\log(Z_G^{\text{OT}}(M))] \leq R_{\text{OT}}(\beta),$$

$$L_{\text{U}}(\beta) \leq \lim_{M \rightarrow \infty} M \text{Var}[\log(Z_G^{\text{U}}(M))] \leq R_{\text{U}}(\beta),$$

where

$$L_{\text{OT}}(\beta) := \frac{e^{2N\beta}}{|\mathcal{X}_G|^2} - 1, R_{\text{OT}}(\beta) := e^{4N\beta} - 1,$$

$$L_{\text{U}}(\beta) := \frac{|\mathcal{X}_G|}{(q + (|\mathcal{X}_G| - q)e^{-8\beta})^2} - 1, R_{\text{U}}(\beta) := e^{8N\beta} - 1.$$

Proof: We have

$$e^{-2N\beta} \leq f_G(x) \leq e^{2N\beta}, \quad (10)$$

$$e^{-2N\beta} \leq \frac{1}{f_G(x)} \leq e^{2N\beta}, \quad (11)$$

and so,

$$e^{2N\beta} \leq Z_G \leq |\mathcal{X}_G| e^{2N\beta}, \quad (12)$$

$$1 \leq \sum_x \frac{1}{f_G(x)} \leq |\mathcal{X}_G| e^{2N\beta}. \quad (13)$$

(The lower bound in (13) is trivially true, and made so to accommodate the case where the grid is of odd size while keeping the derived bounds simple. If the grid is of even size, it can be replaced with $e^{2N\beta}$ —Color the grid in black and white such that no similar colors are adjacent. The lower bound in (12) is when all spins are equal—In fact there are q such configurations and one may replace the lower bound with $q e^{2N\beta}$.) Hence from (8),

$$\begin{aligned} \lim_{M \rightarrow \infty} M \text{Var}[\log(Z_G^{\text{OT}}(M))] &= \frac{Z_G}{|\mathcal{X}_G|^2} \sum_{x \in \mathcal{X}_G} \frac{1}{f_G(x)} - 1 \\ &\stackrel{(13)}{\geq} \frac{Z_G}{|\mathcal{X}_G|^2} - 1 \\ &\stackrel{(12)}{\geq} \frac{e^{2N\beta}}{|\mathcal{X}_G|^2} - 1, \end{aligned}$$

and

$$\begin{aligned} \lim_{M \rightarrow \infty} M \text{Var}[\log(Z_G^{\text{OT}}(M))] &\stackrel{(13)}{\leq} \frac{Z_G}{|\mathcal{X}_G|^2} |\mathcal{X}_G| e^{2N\beta} - 1 \\ &\stackrel{(12)}{\leq} \frac{|\mathcal{X}_G| e^{2N\beta}}{|\mathcal{X}_G|^2} e^{2N\beta} - 1 \\ &= e^{4N\beta} - 1. \end{aligned}$$

From (10) and (11), we also have

$$|\mathcal{X}_G| e^{-2N\beta} \leq Z_G \leq (q + (|\mathcal{X}_G| - q)e^{-8\beta}) e^{2N\beta}, \quad (14)$$

$$e^{4N\beta} \leq \sum_x f_G^2(x) \leq |\mathcal{X}_G| e^{4N\beta}. \quad (15)$$

(In (14) we needed a tighter upper bound than in (12). Instead of trivially replacing the summand with its largest value, we kept its largest q values and replaced the summand's remaining values with its second largest value.) From (9)

$$\begin{aligned} \lim_{M \rightarrow \infty} M \text{Var}[\log(Z_G^{\text{U}}(M))] &= \frac{|\mathcal{X}_G|}{Z_G^2} \sum_{x \in \mathcal{X}_G} f_G^2(x) - 1 \\ &\stackrel{(15)}{\geq} \frac{|\mathcal{X}_G|}{Z_G^2} e^{4N\beta} - 1, \\ &\stackrel{(14)}{\geq} \frac{|\mathcal{X}_G| e^{4N\beta}}{(q + (|\mathcal{X}_G| - q)e^{-8\beta})^2 e^{4N\beta}} - 1, \end{aligned}$$

and

$$\begin{aligned} \lim_{M \rightarrow \infty} M \text{Var}[\log(Z_G^{\text{U}}(M))] &\stackrel{(15)}{\leq} \frac{|\mathcal{X}_G|}{Z_G^2} |\mathcal{X}_G| e^{4N\beta} - 1, \\ &\stackrel{(14)}{\leq} \frac{|\mathcal{X}_G|^2 e^{4N\beta}}{|\mathcal{X}_G|^2 e^{-4N\beta}} - 1. \end{aligned}$$

We remark that the bounds presented in the proposition above (and later in Proposition 2) can be loose for some values of β . However, they suffice to explain the behaviour of the estimators on the primal and dual NFGs. ■

When β is small, say, in the order of N^{-m} for $m > 1$, both upper bounds R_{OT} and R_{U} in the proposition approach zero with increasing N . In this regime both estimators provide good estimates of the partition function, without requiring asymptotically large M .

For large β , however, both estimators are inefficient. In particular, when $\beta > \log q$, the lower bound L_{OT} grows exponentially in N , which requires M to be at least exponential in N in order for the variance to be bounded within a constant. Similarly, when $\beta > \frac{\log q}{8} N$, the lower bound L_{U} also grows exponentially in N , making the uniform estimator inefficient. This is a rather exaggerated value of β , and we refer the reader to [9] for a better discussion on why the uniform estimator is inefficient for large β .

To get a better idea on relative performance between the OT and uniform estimators for large β , note that

$$\lim_{M \rightarrow \infty} M \text{Var}[\log(Z_G^{\text{U}}(M))] \leq |\mathcal{X}_G| - 1,$$

which follows immediately from the fact that $\sum_x f_G^2(x) \leq Z_G^2$. Comparing this upper bound with the lower bound L_{OT} , there exists $\beta_0 := \frac{3}{2} \log(q)$ above which the uniform estimator is more efficient than the OT estimator. This is in fact Theorem 2 of [3] for the model in this work.

On the dual side, we have the following bounds.

Proposition 2. For any integer k , let $A_{k,\beta} := 1 + (k-1)e^{-2\beta}$, and let $r(\beta) := \frac{A_{q,\beta}}{A_{0,\beta}}$. When sampling the dual NFG \mathcal{G}' for the Potts model (with N being an even number),

$$L'_{\text{OT}}(\beta) \leq \lim_{M \rightarrow \infty} M \text{Var}[\log(Z_{\mathcal{G}'}^{\text{OT}}(M))] \leq R'_{\text{OT}}(\beta),$$

$$L'_{\text{U}}(\beta) \leq \lim_{M \rightarrow \infty} M \text{Var}[\log(Z_{\mathcal{G}'}^{\text{U}}(M))] \leq R'_{\text{U}}(\beta),$$

where

$$L'_{\text{OT}}(\beta) := \frac{r^{2N}(\beta)}{|\mathcal{X}_{\mathcal{G}'}|^2} - 1, R'_{\text{OT}}(\beta) := r^{2N}(\beta) - 1,$$

$$L'_{\text{U}}(\beta) := \frac{|\mathcal{X}_{\mathcal{G}'}|}{(q + (|\mathcal{X}_{\mathcal{G}'}| - q)A_{0,\beta})^2} - 1,$$

$$R'_{\text{U}}(\beta) := r^{4N}(\beta) - 1.$$

Proof: We have

$$A_{0,\beta}^{2N} e^{2N\beta} \leq f_{\mathcal{G}'}(x) \leq A_{q,\beta}^{2N} e^{2N\beta}, \quad (16)$$

$$A_{q,\beta}^{-2N} e^{-2N\beta} \leq \frac{1}{f_{\mathcal{G}'}(x)} \leq A_{0,\beta}^{-2N} e^{-2N\beta}. \quad (17)$$

and so,

$$A_{q,\beta}^{2N} e^{2N\beta} \leq Z_{\mathcal{G}'} \leq |\mathcal{X}_{\mathcal{G}'}| A_{q,\beta}^{2N} e^{2N\beta}, \quad (18)$$

$$A_{0,\beta}^{-2N} e^{-2N\beta} \leq \sum_x \frac{1}{f_{\mathcal{G}'}(x)} \leq |\mathcal{X}_{\mathcal{G}'}| A_{0,\beta}^{-2N} e^{-2N\beta}. \quad (19)$$

(The lower bound in (19) is valid since the model is of even

size.) Hence from (8),

$$\begin{aligned} \lim_{M \rightarrow \infty} M\text{Var}[\log(Z_{G'}^{\text{OT}}(M))] &= \frac{Z_{G'}}{|\mathcal{X}_{G'}|^2} \sum_{x \in \mathcal{X}_{G'}} \frac{1}{f_{G'}(x)} - 1 \\ &\stackrel{(19)}{\geq} \frac{Z_{G'}}{|\mathcal{X}_{G'}|^2} A_{0,\beta}^{-2N} e^{-2N\beta} - 1 \\ &\stackrel{(18)}{\geq} \frac{A_{0,\beta}^{-2N} A_{q,\beta}^{2N}}{|\mathcal{X}_{G'}|^2} - 1, \end{aligned}$$

and

$$\begin{aligned} \lim_{M \rightarrow \infty} M\text{Var}[\log(Z_{G'}^{\text{OT}}(M))] &\stackrel{(19)}{\leq} \frac{Z_{G'}}{|\mathcal{X}_{G'}|^2} |\mathcal{X}_{G'}| A_{0,\beta}^{-2N} e^{-2N\beta} - 1 \\ &\stackrel{(18)}{\leq} A_{q,\beta}^{2N} e^{2N\beta} A_{0,\beta}^{-2N} e^{-2N\beta} - 1 \\ &= A_{0,\beta}^{-2N} A_{q,\beta}^{2N} - 1. \end{aligned}$$

From (16) and (17), we also have

$$|\mathcal{X}_{G'}| A_{0,\beta}^{2N} e^{2N\beta} \leq Z_{G'} \leq (q + (|\mathcal{X}_{G'}| - q) A_{0,\beta}) A_{q,\beta}^{2N} e^{2N\beta}, \quad (20)$$

$$A_{q,\beta}^{4N} e^{4N\beta} \leq \sum_x f_{G'}^2(x) \leq |\mathcal{X}_{G'}| A_{q,\beta}^{4N} e^{4N\beta}, \quad (21)$$

where the upper bound in (20) follows from

$$Z_{G'} \leq q A_{q,\beta}^{2N} e^{2N\beta} + (|\mathcal{X}_{G'}| - q) A_{0,\beta}^{2N-1} e^{(2N-1)\beta}.$$

From (9)

$$\begin{aligned} \lim_{M \rightarrow \infty} M\text{Var}[\log(Z_{G'}^{\text{U}}(M))] &= \frac{|\mathcal{X}_{G'}|}{Z_{G'}^2} \sum_{x \in \mathcal{X}_{G'}} f_{G'}^2(x) - 1 \\ &\stackrel{(21)}{\geq} \frac{|\mathcal{X}_{G'}|}{Z_{G'}^2} A_{q,\beta}^{4N} e^{4N\beta} - 1 \\ &\stackrel{(20)}{\geq} \frac{|\mathcal{X}_{G'}| A_{q,\beta}^{4N} e^{4N\beta}}{(q + (|\mathcal{X}_{G'}| - q) A_{0,\beta})^2 A_{q,\beta}^{4N} e^{4N\beta}} - 1 \\ &= \frac{|\mathcal{X}_{G'}|}{(q + (|\mathcal{X}_{G'}| - q) A_{0,\beta})^2} - 1. \end{aligned}$$

and

$$\begin{aligned} \lim_{M \rightarrow \infty} M\text{Var}[\log(Z_{G'}^{\text{U}}(M))] &\stackrel{(21)}{\leq} \frac{|\mathcal{X}_{G'}|}{Z_{G'}^2} |\mathcal{X}_{G'}| A_{q,\beta}^{4N} e^{4N\beta} - 1, \\ &\stackrel{(20)}{\leq} \frac{|\mathcal{X}_{G'}|^2 A_{q,\beta}^{4N} e^{4N\beta}}{|\mathcal{X}_{G'}|^2 A_{0,\beta}^{4N} e^{4N\beta}} - 1, \\ &= A_{0,\beta}^{-4N} A_{q,\beta}^{4N} - 1. \end{aligned}$$

When β is large, namely in the order of $\log(N)$, both upper bounds R'_{OT} and R'_{U} in the proposition approach zero with increasing N . In this regime both estimators provide good estimate of the partition function, without requiring asymptotically large M .

For small β , however, both estimators are inefficient. In particular, for $\beta < \frac{1}{2} \log\left(\frac{2q-1}{q-1}\right)$, the lower bound L'_{OT} grows exponentially in N , which requires M to be at least exponential in N in order for the variance to be bounded

within a constant. Similarly, since $A_{0,\beta}$ approaches zero when β approaches zeros, L'_{U} becomes exponential in N .

Similar to the remark following Proposition 1, comparing $|\mathcal{X}_{G'}|+1$ with the lower bound L'_{OT} , it follows that there exists $\beta'_0 := \frac{1}{2} \log(1 + \frac{q}{q^2-1})$ below which the uniform estimator is more efficient than the OT estimator.

At this end, we have shown that on the dual NFG, the two estimators behave in an opposite trend (in β) to that on the primal NFG. It appears that such a phenomenon may fundamentally be related to a “duality” between “nearly uniform” and “nearly concentrated” distribution. More precisely, when both an NFG and its dual involve only non-negative local functions, they both can be associated with a Boltzmann distribution. If one of the distributions is “nearly uniform”, the other one is necessarily “nearly concentrated”, namely, assigning most of the probability mass to only a few configurations. It is well-known in physics literature that the “near uniformity” and “near concentratedness” correspond respectively to high-temperature and low-temperature systems respectively. It appears that these sampling based estimators usually work well for high-temperature systems and work poorly for low-temperature systems. Taking an NFG to its dual, essentially reverts the “temperature”.

IV. EXPERIMENTS

In this section we provide experimental results for the Potts model with $q = 4$ and grid-size $N = 10 \times 10$. We use the Gibbs sampling algorithm [10] on the primal and dual NFG to obtain samples from p_G and $p_{G'}$, respectively. We estimate the log partition function per site (i.e. $\frac{1}{N} \log(Z)$), where depending on whether the primal or the dual NFG is used, the estimate of the partition function, which depends on the number of samples M , is defined as $\hat{Z}(M) := Z_G^{\text{OT}}(M)$ and $\hat{Z}(M) := q^N Z_{G'}^{\text{OT}}(M)$, respectively. (Similar definitions are used for the uniform estimator.) For any number of samples M , we repeat the experiment 30 times and record the value of $\frac{1}{N} \log(\hat{Z}_i(M))$, $i = 1, \dots, 30$, where for each trial i , the initial configuration is chosen independently and according to the uniform distribution. The “quality” of the estimation at any M is decided based on the standard deviation of the trials from their mean (with respect to the uniform distribution on the set $\{1, \dots, 30\}$).

Figs 4 and 5 show the estimated log partition function per site, i.e., $\log(\hat{Z}(M))/N$, for the low temperature $\beta = 1.2$. Fig. 4 shows the estimation based on the primal NFG using both the uniform estimator (left) and the OT estimator (right). Using up to 10^6 samples, both estimators fail to converge, and so do not provide a good estimation. This can also be seen in the dashed lines in Fig. 6 showing the standard deviation of the uniform estimator (left) and the OT estimator (right), where the standard deviation in both cases remains high. In contrast, Fig. 5 shows fast convergence of the estimators on the dual NFG. The standard deviation of the estimations obtained from the dual NFG is shown using the solid lines in Fig. 6. Fig. 7 shows the standard deviation of the estimations for the high temperature of $\beta = 0.18$. In this case estimations

based on the primal NFG have a lower standard deviation compared to the dual NFG, and so provide a better estimation. In Fig. 8 (a), showing the standard deviation versus β using uniform sampling, one observes the behaviour of the estimator versus β as discussed in Section III.

V. CONCLUDING REMARKS: BEYOND THE POTTS MODEL

This paper shows analytically and experimentally that stochastic estimators of partition functions exhibit opposite trends on NFG representation of a model and its dual. As remarked in Section III, this phenomenon is fundamentally related to a duality between “nearly concentrated” and “nearly uniform” distribution. This understanding allows the results presented above to extend beyond the Potts models. In particular, one may consider two-dimensional nearest neighbor models whose bivariate local function is of the form $h(x, x') := \kappa(x - x')$ for other functions κ . When both κ and $\hat{\kappa}$ are a non-negative real function, the duality between uniformity and concentratedness is expected to hold and such a phenomenon is expected to occur. As an example, consider the “clock model,” which is defined in the same way as the Potts model under the choice

$$\kappa_{\text{clock}}(x) = e^{\beta \cos(2\pi x/q)}, \quad (22)$$

for all $x \in \mathbb{Z}_q$. (Hence, it is within the scope of models of Fig. 2.) From Lemma 1 below, $\hat{\kappa}_{\text{clock}}$ is a positive function, and so it is possible to take the dual NFG route toward estimating its partition function.

Lemma 1. $\hat{\kappa}_{\text{clock}}$ is a positive function.

Proof: For any $x, y \in \mathbb{Z}_q$, let $\chi_y(x) := e^{2\pi\sqrt{-1}xy/q}$. Using Taylor expansion, we have

$$\kappa_c(x) = \sum_{n=0}^{\infty} \frac{\beta^n g_n(x)}{n!},$$

where

$$\begin{aligned} g_n(x) &:= \cos^n(2\pi x/q) = \frac{1}{2^n} (\chi_1(x) + \chi_1(-x))^n \\ &= \frac{1}{2^n} \sum_{l=0}^n \binom{n}{l} \chi_1^{n-l}(x) \chi_1^l(-x) \\ &= \frac{1}{2^n} \sum_{l=0}^n \binom{n}{l} \chi_{\bar{n}-2l}(x), \end{aligned}$$

where for any $m \in \mathbb{Z}$, $\bar{m} \in \mathbb{Z}_q$ is defined as m modulo q . Hence,

$$\hat{g}_n(\chi_k) = \frac{q}{2^n} \sum_{l=0}^n \binom{n}{l} [\bar{n} + k - 2l = 0]$$

is a non-negative function that is upper bounded by q , where for any m , $[m = 0]$ is the indicator function evaluating to one iff $m = 0$. Therefore,

$$\hat{\kappa}_c(\chi_k) = \sum_{n=0}^{\infty} \frac{\beta^n \hat{g}_n(\chi_k)}{n!}$$

is a positive function for $\beta > 0$. (This follows since for any $k \in \mathbb{Z}_q$, $\hat{g}_n(\chi_k)$ cannot be zero for all n . In particular, $\hat{g}_{q-k}(\chi_k) > 0$.) Finally, the series in the RHS is convergent since

$$\sum_{n=0}^{\infty} \frac{\beta^n \hat{g}_n(\chi_k)}{n!} \leq q \sum_{n=0}^{\infty} \frac{\beta^n}{n!} = qe^{\beta}.$$

■

Consider for instance the clock model with $q = 4$. It is not hard to see that p_G in this case is a “concentrated” distribution for low temperatures and an “almost uniform” distribution for high temperatures. From this and the fact that

$$\hat{\kappa}_{\text{clock}}(x) = \begin{cases} e^{\beta} + e^{-\beta} + 2, & x = 0 \\ e^{\beta} - e^{-\beta}, & x \in \{1, 3\} \\ e^{\beta} + e^{-\beta} - 2, & x = 2, \end{cases} \quad (23)$$

one may obtain similar results to Propositions 1 and 2. Simulation results for this model are shown in Fig. 8 (b).

REFERENCES

- [1] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, “A new class of upper bounds on the log partition function,” *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2313–2335, July 2005.
- [2] R. M. Neal, “Probabilistic inference using Markov chain Monte Carlo methods,” Dept. Comp. Science, University of Toronto, Tech. Rep. CRG-TR-93-1, 1993.
- [3] G. Potamianos and J. Goutsias, “Stochastic approximation algorithms for partition function estimation of Gibbs random fields,” *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1948–1965, Nov. 1997.
- [4] M. Molkaie and H.-A. Loeliger, “Partition function of the Ising model via factor graph duality,” in *Int. Symp. Inf. Theory*, Istanbul, Turkey, July 2013, pp. 2304–2308.
- [5] A. Al-Bashabsheh and Y. Mao, “Normal factor graphs and holographic transformations,” *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 752–763, Feb. 2011.
- [6] Y. Ogata and M. Tanemura, “Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure,” *Ann. Inst. Statist. Math.*, vol. 33, no. 1, pp. 315–338, 1980.
- [7] F.-Y. Wu, “The Potts model,” *Reviews of modern physics*, vol. 54, no. 1, pp. 235–268, 1982.
- [8] G. Casella and R. L. Berger, *Statistics Inference*. Pacific Grove, California: Duxbury, 2002.
- [9] D. J. MacKay, “Introduction to monte carlo methods,” in *Learning in graphical models*, 1998, pp. 175–204.
- [10] M. Newman and G. Barkema, *Monte Carlo methods in statistical physics*. Oxford: Oxford University Press, Reprint 2001.

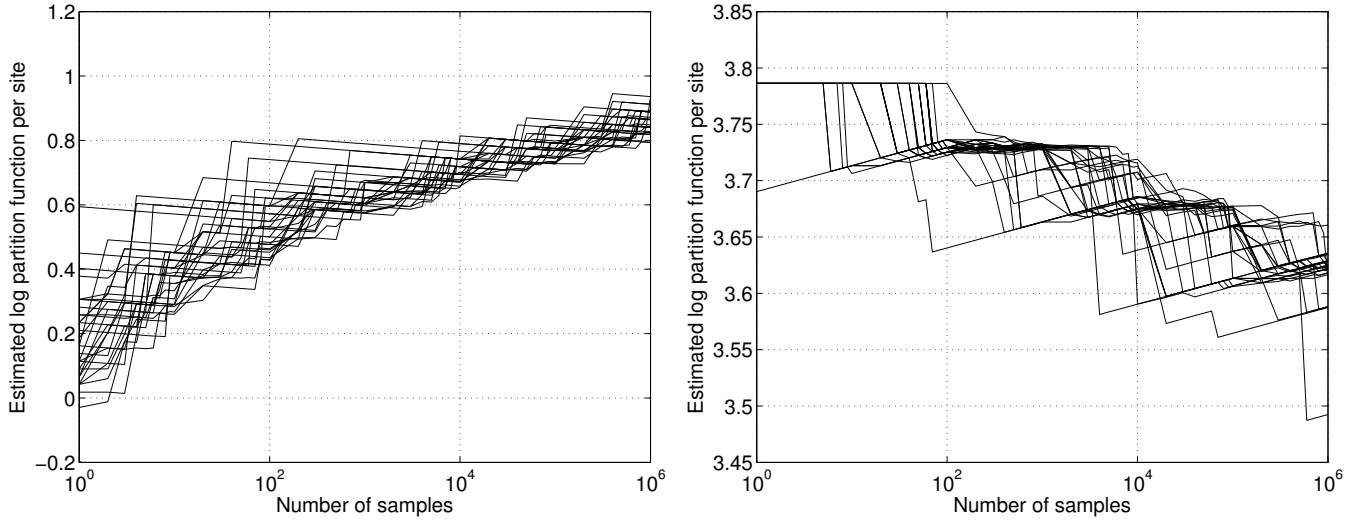


Fig. 4. Potts model at low temperature $\beta = 1.2$ using the primal NFG. The two figures show the estimated log partition function per site versus the number of samples using the uniform estimator (left) and the OT estimator (right), where each line represents a trial.

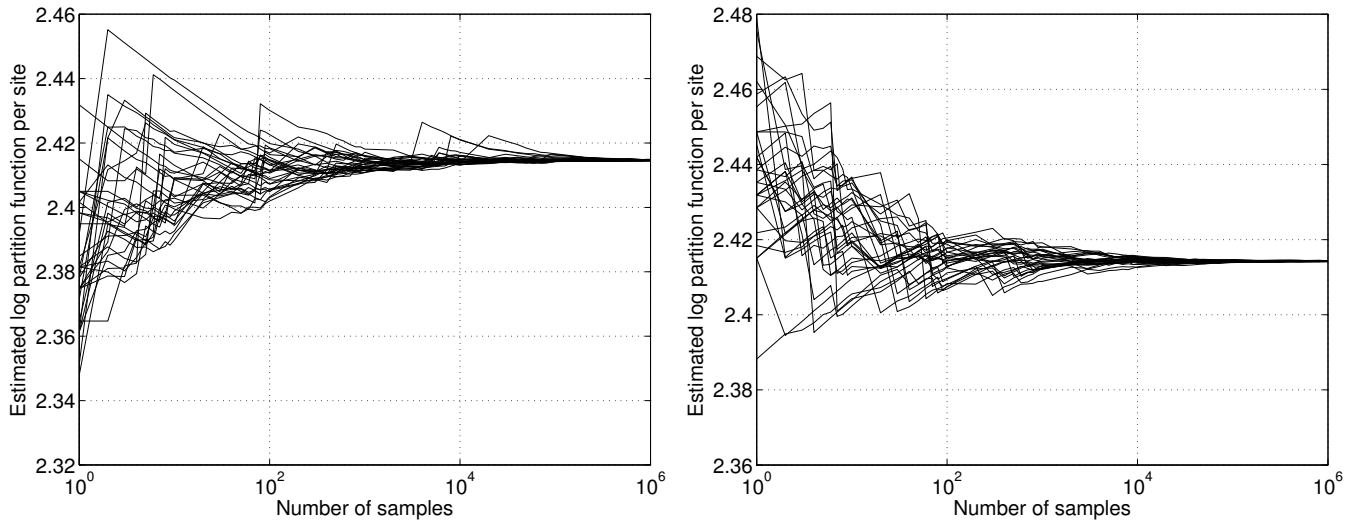


Fig. 5. Potts model at low temperature $\beta = 1.2$ using the dual NFG. The two figures show the estimated log partition function per site versus the number of samples using the uniform estimator (left) and the OT estimator (right), where each line represents a trial.

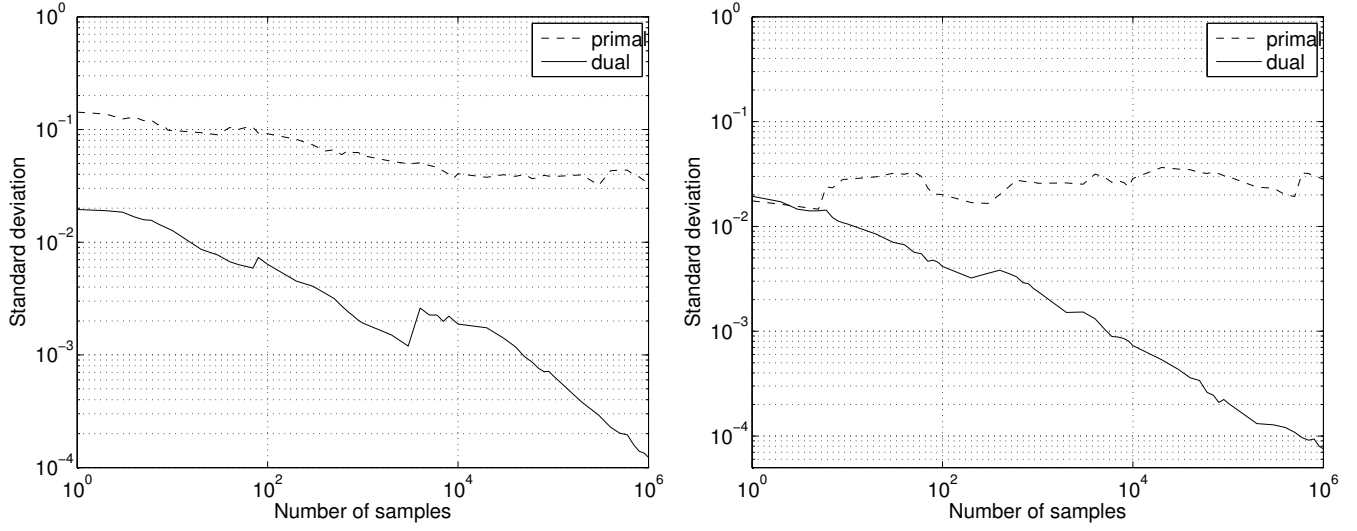


Fig. 6. Potts model at low temperature $\beta = 1.2$, where the standard deviation of the estimated log partition function per site is shown for the uniform (left) and OT (right) estimators based on the primal (dashed line) and dual (solid line) NFGs.

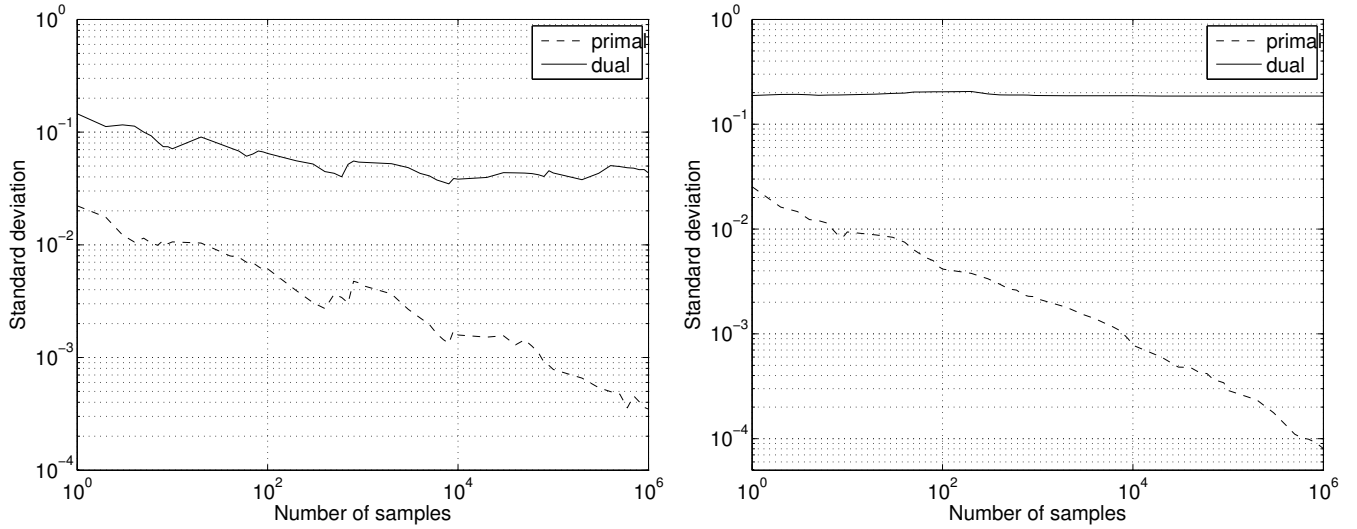
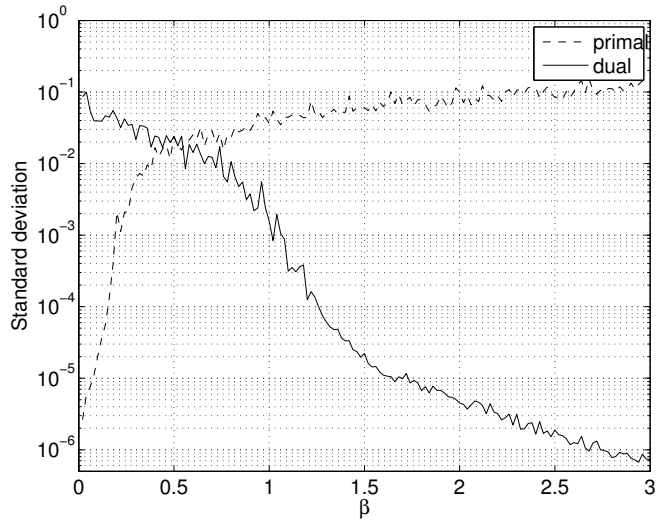
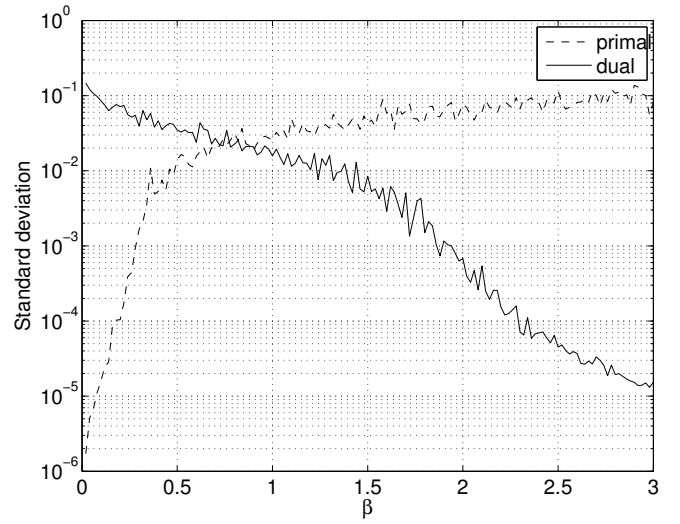


Fig. 7. Potts model at high temperature $\beta = 0.18$, where the standard deviation of the estimated log partition function per site is shown for the uniform (left) and OT (right) estimators based on the primal (dashed line) and dual (solid line) NFGs.



(a) Potts model.



(b) Clock model.

Fig. 8. Standard deviation of the estimated free energy per site versus β using uniform sampling with $M = 10^6$.